1. JP,2002-528797,A
2. JP,2004-508616,A

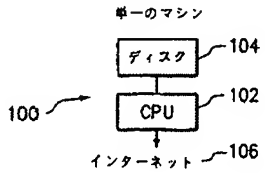
[Translation done.]

DETAILED DESCRIPTION

[Detailed Description of the Invention]
[0001]
[Field of the Invention]
Generally this invention relates to data processing. Especially this invention relates to the method and device which control a computing grid.
[0002]
[Problem(s) to be Solved by the

Drawing selection

drawing 1 A



[Translation done.]

DETAILED DESCRIPTION

[Detailed Description of the Invention]
[0001]
[Field of the Invention]
Generally this invention relates to data processing. Especially this invention relates to the method and device which control a computing grid.
[0002]
[Problem(s) to be Solved by the Invention]
The builder of today's website and other computer systems has many interesting system planning problems. A capacity planning, site availability, and the safety of the site are contained in these problems. In order to attain these aims, it is required to discover and employ the staff who received training in which the design and management of a site which may be greatly complicated potentially are possible. It turns out that it is difficult to discover and employ such a staff for many organizations since the design of a big site, construction, and management are not the prime enterprises in many cases.
[0003]
The company website of the third party site arranged at the place where other websites of other companies are the same as one method was adopted. Such an outsourcing institution can be used from companies, such as Exodus, AboveNet, and GlobalCenter, now. The physical space, the redundant network, and electric generating facilities which many customers share are given by these institutions.
[0004]
By adoption of an outsourcing website, although establishment of a website and the burden of maintenance decrease greatly, it does not become removing all the problems relevant to maintenance of a website from a company. The company has to perform many work about the computing structure base between construction of the institution, management, and increase.

The info technology administrator of the company employed in such an institution is responsible about manual selection of the arithmetic unit in institutions, installation, composition, and maintenance. The administrator has to tackle a problem with difficult resource planning, handling peak capacity, etc. Especially the administrator needs to predict resource demand and a demand resource from an outsourcing company, in order to cope with demand. As a relaxation measure over unexpected peak demand, more than many administrators need, they secure capacity sufficient by requiring many resources substantially. Though regrettable, it will become great [intact capacity] by this, and the overhead charge of the company for adopting a website will increase.

[0005]

Since the same manual management treatment that is easy to be mistaken is needed with growth even if an outsourcing company also provides a perfect calculation facility including a server, software, and a power plant, expansion and growth of institutions are not easy for an outsourcing company. The problem remains with the capacity planning to unexpected peak demand. In this case, an outsourcing company may maintain most quantity of intact capacity.

[0006]

The necessary condition of the website which an outsourcing company manages has often differed. For example, the capability for managing and controlling the website independently is needed in a certain company. The security of the specific kind into which the website is made to separate from all the sites of the others arranged in the outsourcing company, or a level is needed in both other companies. As another example, the positive connection with the corporate intranet arranged at somewhere is needed in a certain company..

[0007]

Various websites differ in internal topology. A certain site only comprises a web server sequence which load balance was able to take by the web load balancer. Suitable load balancers are Cisco Systems, Local Director of Inc., BigIP of F5Labs, Web Director of Aleton, etc. Since multilayered constitution of other sites is carried out, the web server sequence can cope with a hypertext protocol (HTTP) demand by this, but the great portion of application logic is carried out in another application server. These application servers may have to be again connected to the layer of a database server.

[0008]

Some of such different structure scenarios are shown in drawing 1 A, drawing 1 B, and drawing 1 C. Drawing 1 A is a block diagram of a simple website, and comprises the single computing element or the machine 100 containing CPU102 and the disk 104. The machine 100 is connected to the packet-switched data networks 106 of the worldwide scale known as the Internet, or other networks. The machine 100 may be accommodated in the same position service of the type mentioned above.

[0009]
Drawing 1 B is a block diagram of the one-layer web server firm 110 containing two or more web servers WSA, WSB, and WSC. Each web server is connected to the load balancer 112 connected to the Internet 106. A load balancer divides the traffic between servers and maintains the processing load which the balance of each server was able to take. It may be connected to this, including the firewall for also protecting the load balancer 112 from the traffic to which the web server is not permitted.
[0010]
Drawing 1 C shows the three-layer server farm 120 containing layers, such as the web server W1 and W2, layers, such as the application server A1 and A2, and layers, such as the database server D1 and D2. A web server is provided in order to cope with an HTTP request. An application server performs the great portion of application logic. A database server performs database management system (DBMS) software.
[0011]
The topology of the kind of website which needs to be constituted is diversified, and since the necessary condition of an applicable company is changing, it is thought that the only method of constituting a large-scale website is customizing each site physically. Many organizations are tackling the same problem individually, respectively, and are customizing each website from zero. A lot of [it is inefficient and] same work in a different company will produce this.
[0012]
Another problems of the conventional method are a resource and a capacity planning. A website is a different day or time to differ of the day, and receives the traffic of a dramatically different level. In peak traffic time, the hardware or software of a website may be unable to answer a demand by within a time [suitable for an overload]. In other time, there is excessive capacity in the hardware or software of a website, and it is not fully used. It is a difficult problem to find the balance in having the sufficient hardware and software coping with peak traffic, without undertaking excessive cost in the conventional method, or becoming superfluous capacity. Many websites cannot find suitable balance but are chronically afflicted by too little capacity or superfluous capacity.
[0013]
Another problem is failure caused by the human error. The serious potential disaster which exists in the present method of using the server farm by which manual composition was carried out, It is that a server farm will malfunction by a human error when a new server is constituted in a live server, and service to the user of a website may be lost by this.
[0014]
As soon as there is a demand without needing custom-made composition in this field based on the above, the method and device which provide an easily extensible promptly computing

system and which have been improved are clearly required.
[0015]
In order to clarify change of a traffic throughput, the computing system which supports generation of the separation node of a large number in which extension or reduction is possible if needed, respectively is also required.
[0016]
The method and device which control such an extensible computing system and its composition separation node are also required. It will become clear [other necessity] from the disclosure shown here.
[0017]
[Description of the Invention]
According to one mode of this invention, the above-mentioned necessity and other necessity which becomes clear by the following explanation, Based on a large-scale computing structure ("computing grid"), it is dramatically extensible and can be reached by the method and device which are very easy to use, and control and manage a positive data-processing site. A computing grid is constituted physically and is logically divided to various organizations according to a demand after that. The computing grid includes many KOMPYUTINGUGU elements in the emergency connected to 1 or two or more VLAN switches and 1, or two or more storage area network (SAN) switches. It may be connected to a SAN switch and two or more memory storage may be selectively connected to 1 or two or more computing elements via suitable change logic and command. One port of a VLAN switch is connected to external networks, such as the Internet. Surveillance, a layer, a machine, or a process is connected to a VLAN switch and a SAN switch.
[0018]
Introduction, all the memory storage, and computing elements are assigned to an idol pool. Under programmed control, surveillance constitutes dynamically so that the port of a VLAN switch and a SAN switch may be connected to 1 or two or more computing elements, and memory storage. As a result, such an element and a device are logically removed from an idol pool, and become some of 1, two or more virtual server farms (VSF), or instant data centers (IDC). In order to perform bootstrap operation and generation execution, each VSF computing element is turned to the memory storage containing the boot image which can use a computing element, or is associated.
[0019]
According to one mode of this invention, a surveillance layer is control plane which comprises the control mechanism hierarchy containing 1 or two or more master controlling process mechanisms by which the communication interface was carried out to 1 or two or more slave controlling process mechanisms. One or two or more master controlling process mechanisms

are based on loading of a slave controlling process mechanism, are assigned and assigned and cancel a slave controlling process mechanism. By choosing the subset of processing and a memory resource, 1 or two or more master controlling process mechanisms are supported in a slave controlling process mechanism so that IDC may be established. One or two or more master controlling process mechanisms perform the periodic medical checkup of a slave controlling process mechanism. There is no response or the slave control mechanism which carried out abnormal termination is rebooted. It is started and another slave control mechanism serves as instead of [of the slave control mechanism which cannot be resumed]. A slave control mechanism performs the periodic medical checkup of a master control mechanism. If a master slave controlling process mechanism carries out abnormal termination, a slave controlling process mechanism will be chosen, and it will become a new master controlling process mechanism, and will become instead of [of the master controlling process mechanism ended above].

[0020]

At the time of customization of each site, the difficult economy of scale is obtained by constituting a computing grid physically once and assigning the portion of a computing grid to certain and dynamically various organizations according to a demand.

[0021]

In the attached drawing, this invention is illustrated as an example rather than is limited, and the same reference number shows the same element in it.

[0022]

[Embodiment of the invention]

In the following explanation, in order to let you understand this invention thoroughly for the purpose of explanation, much specific details are described. However, it will become clear to a person skilled in the art that this invention is carried out without these specific details. In other examples, in order to avoid that this invention becomes unclear superfluously, a known structure and device are shown by the block diagram.

[0023]

Virtual server farm (VSF)

According to one example, a large-scale computing structure ("computing grid") is established. A computing grid may be constituted once physically and may be logically divided according to a demand after that. A part of computing grid is assigned to two or more companies or each of an organization. The logic part of the computing grid of each organization is called a virtual server farm (VSF). Each organization maintains the management control which the VSF became independent of. Each VSF can change dynamically the number of CPUs, a storage capacity and a disk, and network band width based on the real-time demand given to a server farm or other elements. Although VSF is altogether generated logically from the same physical

computing grid, each VSF is protected from VSF of all other organizations. VSF can be conversely connected to intranet by using an individual dedicated line or a virtual private network (VPN), without exposing intranet to VSF of other organizations.

[0024]

Although the organization can perform completeness (for example, superuser or route) management access to a computer and all the traffic of the Local Area Network (LAN) where these computers were connected can be observed, Only the portion, i.e., the data in VSF, and computing element of the computing grid assigned to it can be accessed. According to one example, this becomes possible by using the dynamic firewall method which the safety clearance of VSF extends and reduces dynamically. Each VSF can be used and the contents and application of an organization which can be accessed via the Internet, intranet, or an extra network can be adopted.

[0025]

A computing element, its related networking, and the composition and control of a storage element are performed by the surveillance which cannot do direct access depending on any of the computing element in a computing grid they are. For convenience, in this document, generally surveillance may be called control plane and may comprise a network of 1, two or more processors, or a processor. Surveillance may comprise a supervisor, a controller, etc. Other methods can also be used so that it may explain here.

[0026]

For example by the inside of a network, or other means, control plane is on a full independence set of the computing element assigned to the purposes of surveillance, such as 1 or two or more servers by which interconnection is carried out, and is carried out. Control plane performs a control action to computing, the networking, and the storage element of a computing grid via the special control port or interface of the networking of a grid, and a storage element. Control plane gives a physical interface to the change element of a system, supervises the load of the computing element in a system, and gives a management function using a graphical user interface or other suitable user interfaces.

[0027]

To the computer in a computer grid (and specific VSF), the computer used for carrying out control plane is invisible logically, and via the element in a computer grid, Or from an external computer, it is never attacked or is not destroyed. Only control plane has a physical-connection part to the control ports of the apparatus in a computer grid, and this controls the membership in specific VSF. Since the apparatus in computing can be constituted via these special control ports, The computing element in a computing grid cannot change the safety clearance, or cannot perform access to the memory or computing apparatus which is not accepted.

[0028]

Therefore, an organization can be interlocked with the private server farm dynamically made from the large-scale share computing infrastructure, i.e., the computing equipment which seems to have comprised a computing grid, by VSF. The control plane connected with the computing architecture explained here gives the private server firm with which the privacy and integrity are protected by the access controller carried out in the hardware of the apparatus of a computing grid.

[0029]

Control plane controls the internal topology of each VSF. The control plane can take the basic interconnection of the computer explained here, a network switch, and a memory network switch, and can create various server farm composition using these. Although not limited to these, the monolayer web server firm pretreated by the load balancer and multilayered constitution are included, and a web server communicates with an application server, and an application server communicates with a database server. Various loads balancing, multilayering, and firewall composition are possible.

[0030]

Computing grid

The computing grid can exist in a single place and a broad field can be made to distribute it. First, this book explains the computer grid in the network of the size of the single building which comprises only local-area art. Next, this book explains the case where a computer grid is distributed on a wide area network (WAN).

[0031]

Drawing 2 is a block diagram showing one composition of the extensible computing system 200 containing the local computing grid 208. In this book, as soon as a system is flexible, it is extensible as generally as "extensible" and there is a demand, it means having the capability to give the calculation power in which it was made to fall or go up to a specific company or user. the local computing grid 208 -- much computing element CPU1, CPU2, and ... CPUn is comprised. 10,000 or more computing elements exist in an example. Since these computing elements do not include the state information for every long-term element or do not save it, it may constitute without the perpetuity of a local disk etc., or nonvolatile storage. Instead, apart from a computing element, all the long-term state information, two or more disks, the disk 1 and the disk 2 which are connected to a computing element via a storage area network (SAN) including 1 or the two or more SAN switches 202, and ... it is saved on the disk n. The example of the suitable SAN switch is sold from Brocade and Excel.

[0032]

Interconnection of all the computing elements is carried out via 1 or the two or more VLAN switches 204 which are divided into virtual LAN (VLAN). The VLAN switch 204 is connected to the Internet 106. Generally, the computing element includes one or two network interfaces

which were connected to the VLAN switch. For convenience, in drawing 2, although all the nodes have two network interfaces, many [that it is less than this or] nodes also have a network interface. Many manufacture supply origin provides now the switch which supports a VLAN function. For example, it is more nearly available than Cisco Systems, Inc, and Xtreme Networks in a suitable VLAN switch. Similarly, and a fiber channel switch, SCSI versus Fibre Channel bridging apparatus, and network attaching DOSUTO rage (NAS) apparatus are contained in this. [the available products for constituting SAN]

[0033]

the control plane 206 -- a SAN control route, a CPU control course, and a VLAN control route -- the SAN switch 202, CPU1, CPU2, and ... it is connected to CPUn and the VLAN switch 204, respectively.

[0034]

Each VSF comprises the subset of available memory storage on SAN connected to a set of VLAN, a set of the computing element attached to VLAN, and a set of a computing element. The subset of available storage is called a SAN zone on SAN, and this is protected from access from the computing element which are a part of other SAN zones by SAN hardware. A one customer or end user is prevented from using suitably VLAN which gives an immalleable port identifier, and accessing other customers or the VSF resource of an end user.

[0035]

Drawing 3 is a block diagram of the typical virtual server farm which makes a SAN zone the special feature. Two or more web server WS1, WS2, etc. are connected to load balancer (LB) / firewall 302 by the 1st VLAN (VLAN1). The 2nd VLAN (VLAN2) connects the Internet 106 to load balancer (LB) / firewall 302. Each web server can be chosen from CPU1, CPU2, etc. using the mechanism explained later. The web server is connected to the SAN zone 304, and this is connected to 1 or the two or more memory storage 306a and 306b.

[0036]

In a certain time, the computing element in computing grids, such as CPU1 of drawing 2, is only connected to the set of VLAN, and the SAN zone relevant to single VSF. Usually, VSF is not shared among different organizations. A set of VLAN relevant to the subset of the storage on SAN belonging to a single SAN zone and it and the computing element on these VLAN(s) specify VSF.

[0037]

By controlling the membership of VLAN, and the membership of a SAN zone, control plane carries out logic partitioning of the computing grid to many VSF(s). The member of one VSF cannot access other computings or memory resources of VSF. such an access restriction -- a VLAN switch -- and it is made to perform with the port level-access-control mechanism (for example, zoning) of SAN hardware called edge apparatus, such as a fiber channel switch and

SCSI versus Fibre Channel bridging hardware Since it is not physically connected to the control ports or the interface of a VSAN switch and a SAN switch, the computing element which forms a part of computing grid cannot control the membership of VLAN or a SAN zone. Therefore, the computing element of a computing grid cannot access the computing element which is not arranged at VSF containing these.

[0038]

Only the computing element which performs control plane is physically connected to the control ports or the interface of apparatus in a grid. The apparatus (a computer, a SAN switch, and VLAN switch) of a computing grid is only constituted by these control ports or interfaces. Thereby, a means dramatically stable although it was simple to divide a computing grid into many VSF(s) dynamically is obtained.

[0039]

Each computing element in VSF is as exchangeable as other computing elements. The number of the computing element relevant to a certain VSF, VLAN, and SAN zones will change, if time passes under control of control plane.

[0040]

In one example, the computing grid includes the idol pool which comprises many spare computing elements. The computing element from an idol pool may be assigned to specific VSF for the reasons of the management to failure of the increase in CPU, memory space available at the VSF, or the specific computing element in VSF, etc. When the computing element is constituted as a web server, an idol pool changes or functions as big "shock absorber" to a "letter of burst" web traffic load, and a related peak processing load.

[0041]

Since an idol pool is shared among the organizations where a large number differ, in order not to say that a single organization has to pay the expense of the whole idol pool, economy of scale is obtained. Since a different organization can acquire a computing element from an idol pool in the time of the day to differ if needed, it becomes possible to reduce, when each VSF is expanded when required, and traffic settles in the usual state. As for an idol pool, when the organization where a large number differ continues reaching a peak simultaneously and the capacity of an idol pool may be used up by it, it is possible to make it increase by adding further many CPUs and storage elements to it (extendibility). The capacity of the idol pool is designed in the usual state reduce greatly the probability that another computing element cannot be acquired from an idol pool when specific VSF is required.

[0042]

Drawing 4 A, drawing 4 B, drawing 4 C, and drawing 4 D are the block diagrams showing a continuous process when taking a computing element in and out of an idol pool. With reference to drawing 4 A, control plane should connect the element of the computing grid to

the beginning logically at the 1st and 2nd VSF(s) of a label called VSF1 and VSF2. The idol pool 400 comprises two or more CPU402, and label attachment of one of them is carried out with CPUX. In drawing 4 B, another computing element is needed by VSF1. Therefore, control plane moves CPUX to VSF1 from the idol pool 400, as the course 404 shows.

[0043]

In drawing 4 C, since CPUX is not required for VSF1 any longer, control plane returns CPUX to the idol pool 400 from VSF1. In drawing 4 D, another computing element is needed by VSF2. Therefore, control plane moves CPUX to VSF2 from the idol pool 400. Therefore, when time passes and the state of traffic changes, a single computing element will belong to an idol pool (drawing 4), and it will be assigned to specific VSF (drawing 4 B), and will be returned to an idol pool (drawing 4 C), and will belong to another VSF (drawing 4 D).

[0044]

In each of these stages, control plane constitutes the LAN switch and SAN switch relevant to the computing element which becomes a part of VLAN relevant to specific VSF (or idol pool), and SAN zone. setting between each transition according to one example -- a computing element -- power down -- or it is rebooted. If the power supply of a computing element is switched on again, a computing element will look at the portion from which the memory zone of SAN differs. Especially a computing element looks at the portion of the memory zone on SAN including the image of operating systems (for example, Linux, NT, Solaris, etc.) which can be started. A memory zone contains a data part peculiar to each organization again (for example, the file relevant to a web server, a database partition, etc.). Since a computing element is a part of another VLAN which is a part of another VLAN set of VSF, it can access CPU relevant to VLAN of VSF of the destination, SAN memory storage, and NAS apparatus again.

[0045]

In the suitable example, the memory zone includes two or more logic detail designs relevant to the role assumed by a computing element defined beforehand. Introduction and neither of the computing elements is assigned to specific roles and tasks, such as a web server, an application server, and a database server. The role of a computing element is acquired from either of two or more saved detail designs which were defined beforehand, and each of such a detail design defines the boot image of the computing element relevant to the role. A detail design is saved by the file which makes a boot image position related with a role, a database table, or other preservation format.

[0046]

Therefore, movement of CPUX in drawing 4 A, drawing 4 B, drawing 4 C, and drawing 4 D is logical, is not physical, and is performed by reconstructing a VLAN switch and a SAN zone under control of control plane. Originally it can substitute for each computing element in a computing grid first, and only after being connected to a virtual server farm and loading

software from a boot image, a specific processing role is assumed. As for neither of the computing elements, a specific role or tasks, such as a web server, an application server, and a database server, are assigned. The role of a computing element is acquired from either of two or more saved detail designs which were defined beforehand, and each of these detail designs relates to the role, and defines the boot image of the computer element relevant to a role.

[0047]

Since long-term state information is not saved to specific computing elements (local disk etc.), between different VSF(s), a node is easily movable and can perform completely different OS and application software. Thereby, the plan was made or it becomes easier to exchange a computing element in the case of the down time which is not planned.

[0048]

The specific computing element can perform a different role, when taking in and out of various VSF(s). For example, a computing element will become a database server, a web load balancer, a firewall, etc., if it operates as a web server and is made to move to another VSF in a certain VSF. In different VSF, the operating system with which Linux and NT differ from Solaris etc. can also be started and performed continuously. Therefore, it can substitute for each computing element in a computing grid, and there is no fixed role assigned to it. Therefore, the whole reserve capacity of a computing grid can be used and a certain service which which VSF needs can be provided. Since the number of the backup servers which can provide the same service that each server which performs specific service by this has is set to thousands, the availability and reliability of service which single VSF provides become very high.

[0049]

The dynamic load balancing characteristic and high processor availability are obtained by the high reserve capacity of a computing grid. This capability becomes possible in the most important combination of the diskless computing element which interconnection is carried out via VLAN, and is connected to the zone of memory storage which can be constituted via SAN, and is altogether controlled by control plane in real time. Each computing element can operate in the role of which required server in VSF, and can be connected to which logic partitioning of which disk in SAN. By a grid, when the further computing power and disk storage capacity are required, a computing element or disk storage is manually added to an idol pool, but this will decrease, if time passes and VSF service is provided for further many organizations. It is not necessary to intervene in increasing the number of CPUs, a network and disk processing capability, and the memory storage that can be used by VSF manually. All these resources are assigned by control plane from the network which can be used to CPU and an idol pool whenever there is a demand, and a disk resource.

[0050]
Specific VSF is not reconstructed manually. Only the computing element of an idol pool is manually reconstructed by the computing grid. As a result, the serious potential obstacle which exists in the server farm which comprised present hand control is removed. A server farm malfunctions by the human error at the time of constituting a new server in a live server farm, and most possibilities that service to the user of the website will be lost as a result disappear.
[0051]
Since control plane copies the data saved at the memory storage attached to SAN again, service into which portion of a system is not lost by failure of a specific storage element. Since every computing element can be attached to which memory partition by removing a long-term memory device from a computing device by using SAN and giving redundant memory and a computing element, high availability is acquired.
[0052]
The detailed example of the addition of the processor to establishment of a virtual server farm, and it, and removal of the processor from it
Drawing 5 is a block diagram of the computing grid by an example, and a control plane mechanism. The detailed process which can be used for referring to drawing 5, creating VSF below, and adding a node to it, and removing a node from it is explained.
[0053]
Drawing 5 shows the computing element 502 containing computer A-G connected to the VLAN capable switch 504. The VLAN switch 504 is connected to the Internet 106, and the VLAN switch has the port V1, V2, etc. Computer A-G is further connected to the SAN switch 506, and this is connected to two or more memory storage or disks D1-D5. The SAN switch 506 has the port S1, S2, etc. The communication interface of the control plane mechanism 508 is carried out to the SAN switch 506 and the VLAN switch 504 by the control route and the data path. The control plane can transmit control commands to these devices via control ports.
[0054]
For convenience, the number of the computing elements of drawing 5 has decreased. Actually, many computers (thousands or more [for example,]) and the memory storage of the same number form the computing grid. In such a big structure, interconnection of many SAN switches is carried out, a mesh is formed, and interconnection of the VLAN switch is carried out and it forms a VLAN mesh. However, in order to make it intelligible, drawing 5 shows the single SAN switch and the single VLAN switch.
[0055]
First, all the computer A-G is assigned to the idol pool until control plane receives the preparing request of VSF. All the ports of the VLAN switch are assigned to specific VLAN by which label attachment is carried out with VLANI (for idol zones). It shall be required that

control plane should constitute VSF and one the load balancer / firewall, and two web servers which were connected to the memory storage on SAN shall be included. The demand to control plane is received via management interfaces or other computing elements.
[0056]
It responds to it, and control plane specifies or assigns CPUA as a load balancer/a firewall, and CPUB and CPUC are assigned as a web server. CPUA is logically put on the SAN zone 1, and is turned to the partition on a disk including a load balancer / firewall software for exclusive use which can be started. The word of "being turned" is used for convenience and it means that sufficient information for CPUA to obtain or discover the suitable software which needs to be operated by what kind of means is given to CPUA. By arranging CPUA in the SAN zone 1, it enables CPUA to obtain a resource from the disk controlled by SAN of the SAN zone.
[0057]
A load balancer is constituted by control plane in order to know about CPUB and CPUC as two web servers which should carry out load balance. Firewall composition protects CPUB and CPUC from access which is not accepted from the Internet 106. An operating system with specific CPUB and CPUC. It is turned to the disk partition on SAN containing the OS image for starting for (for example, Solaris, Linux, NT, etc.) and web server application software (for example, Apache). A VLAN switch is constituted so that the ports v1 and v2 may be arranged to VLAN1 and the port v3, v4, v5, v6, and v7 may be arranged to VLAN2. Control plane constitutes the SAN switch 506 and arranges the fiber channel switch port s1, s2, s3, and s8 in the SAN zone 1.
[0058]
It is explained here what kind of meaning CPU is turned to a disk drive specific [how], and, as for this, starting and the shared access to disk data have.
[0059]
Drawing 6 is a block diagram showing the result of the logical connection of the computing element collectively called VSF1. Disk drive DD1 is chosen from the memory storage D1, D2, etc. An invocation command will be given to CPU A, B, and C if the logical structure shown in drawing 6 is acquired. According to it, CPUA serves as an exclusive load balancer / firewall computing element, and CPUB and CPUC serve as a web server.
[0060]
Now, control plane should judge that another web server was required in VSF1 for the rule based on a plan. For example, this happens by the increase in the demand to a web server, and it becomes possible to add at least three web servers to VSF1 by a customer's plan. Or it is because it added by control mechanisms, such as a possible exclusive web page etc. of the organization which owns or manages VSF wanting another server, and adding a server to the VSF further.

[0061]
According to it, it is determined that control plane adds CPUD to VSF1. Therefore, control plane is adding the ports v8 and v9 to VLAN2, and adds CPUD to VLAN2. The SAN port s4 of CPUD is added to the SAN zone 1. CPUD is turned to the portion of the SAN memory storage started and performed as a web server which can be started. Read-only access of the CPUD is carried out at the shared data on SAN which comprises again the contents of a web page, the server script which can be performed, etc. Thus, the web demand turned to the server farm can be coped with so that CPUB and CPUC may satisfy a demand. Control plane constitutes a load balancer (CPUA) so that CPUD may be included as some server sets by which the load balancing is carried out.
[0062]
Next, CPUD was started and the size of VSF increased to three web servers and one load balancer. Drawing 7 shows the logical connection nature obtained as a result.
[0063]
Control plane shall receive the demand which creates another VSF which needs two web servers and one load balancer by the name of VSF2. Control plane assigns CPUE so that it may become a load balancer/firewall, and it assigns CPUF and CPUG so that it may become a web server. CPUE is constituted in order to know about CPUF and CPUG as two computing elements which carry out a load balancing again.
[0064]
In order to carry out this composition, control plane constitutes the VLAN switch 504 so that the ports v10 and v11 may be included in VLAN1 (it is got blocked and the Internet 106 is accessed) and the port v12, v13, v14, and v15 may be contained in VLAN3. Similarly, the SAN switch 506 is constituted so that the SAN port s6, s7, and s9 may be contained in the SAN zone 2. This SAN zone contains memory storage including software required for making it perform as a web server which uses the share read-only disk part contained in the disk D2 of a SAN zone in CPUF and CPUG by making CPUE into a load balancer.
[0065]
Drawing 8 is a block diagram of the logical connection nature obtained as a result. Although two VSF(s) (VSF1, VSF2) share the same physical VLAN switch and a SAN switch, two VSF (s) are divided logically. CPU
It is only that the company which owns or manages the user or VSF1 which accesses B, C, and D can access CPU and storage of VSF1. Such a user cannot access CPU or storage of VSF2. This cannot perform such access for separate VLAN on the only share segment (VLAN1), the combination of two firewalls, and a different SAN zone where two VSF(s) are constituted.
[0066]

It shall be judged that the control plane can return VSF1 to two web servers. This is because the transient rise of the load of VSF1 fell or other administrative actions were taken. The shut [control plane / with a special command including the power OFF of CPU / CPUD] according to it. If shut [CPU], control plane will demount the ports v8 and v9 from VLAN2, or will remove from the SAN zone 1 with the SAN port s4. The port s4 is arranged in an idol SAN zone. An idol SAN zone is specified as the SAN (for idols) zone I, or the zone 0, for example.

[0067]

Then, it is determined that control plane adds another node to VSF2. This is for according to the reason of others [**** / that the load of the web server in VSF2 rises temporarily]. Therefore, control plane determines to arrange CPUD to VSF2, as the dashed line course 802 shows. Therefore, a VLAN switch is constituted so that the ports v8 and v9 may be included in VLAN3 and the SAN port s4 may be included in the SAN zone 2. CPUD is turned to the memory portion of the disk unit 2 including the image for starting of OS required for the server of VSF2, and web server software. Read-only access to the data of the file system with which other web servers of VSF2 share CPUD is permitted. A power supply is switched on again, and CPUD is performed as a web server in VSF2 by which the load balancing was carried out, and is not accessed to CPU attached to the data or VLAN2 in the SAN zone 1. CPUD in particular cannot be accessed at the time of the first stage which was a part of VSF1 at the element of VSF1, either.

[0068]

In this composition, the safety clearance performed by CPUE was dynamically extended until CPUD was included. Therefore, the dynamic firewall automatically adjusted so that the computing element added or removed by VSF may be appropriately protected according to an example is provided.

[0069]

The example explained the SAN zoning based on a port for explanation. The SAN zoning of other kinds can also be used. For example, LUN level SAN zoning may be used and a SAN zone may be created based on the amount of logic in a disk array. Example products suitable for LUN level SAN zoning are Volume Logics Product of EMC Corporation.

[0070]

A disk unit on SAN

There are some methods of turning CPU to the specific device on SAN in order to call it access to the disk storage which has the information about where starting or other disk storage which need to be shared, a boot program, and data are found.

[0071]

In one method, the SCSI interface of the SCSI versus Fibre Channel bridging apparatus attached to the computing element and a local disk is established. By determining the course

from the SCSI port to the suitable apparatus of Fibre Channel SAN, as a computer accesses the SCSI equipment attached locally, it can access the memory storage on Fibre Channel SAN. Therefore, as software, such as starting software, carries out boot-off of the SCSI equipment attached locally, it carries out boot-off of the disk unit on SAN simply.

[0072]

An option is booting ROM and OS software of a node which have a device driver which Fibre-Channel-interfaces and is related, and make a Fibre Channel interface usable as boot apparatus.

[0073]

In other methods, although it becomes SCSI or an IDE apparatus controller, it has an interface card (for example, a PCI bus or S bus) which communicates on SAN and accesses a disk. Operating systems, such as Solaris, make full offer of the usable diskless boot function by this method.

[0074]

Usually, there are two kinds of SAN disk apparatus relevant to a certain node. One kind is not logically shared with other computing elements, but constitutes the thing containing the OS image which can be started, a local configuration file, etc. which is usually a root partition for every node. This is equivalent to the root file system on a Unix (registered trademark) system.

[0075]

The disk of the 2nd kind is share storage with other nodes. A shared kind changes with needs of the node which accesses OS software and share storage which are performed on CPU. When OS provides the cluster file system which makes possible reading/writing access of a share disk partition among many nodes, a shared disk is mounted as such a cluster file system. Similarly, in order to perform simultaneous reading / writing access to a shared disk, database software, such as an oracle parallel server which enables execution of many nodes within a cluster, may be used for a system. In such a case, the shared disk is already designed in a fundamental OS and application software.

[0076]

Since OS and related application cannot manage the disk apparatus shared with other nodes in the case of the operating system in which such shared access is impossible, a shared disk can be mounted as read-only apparatus. What is necessary is just to carry out read-only access to a web related file in the case of many Web site applications. For example, in the case of a Unix (registered trademark) system, a specific file system may be mounted as read-only.

[0077]

Multi-switch computing grid

The composition explained above in relation to drawing 5 by carrying out interconnection of

two or more VLAN switches, and forming a big exchange VLAN structure, And by carrying out interconnection of many SAN switches, and constituting a big exchange SAN mesh, it is extensible to many computings and memory nodes. In this case, as for the computing grid, a SAN/VLAN exchange mesh has the architecture generally shown in drawing 5 except for the thing of CPU and memory storage for which many ports are included dramatically. The computing element of a large number which perform control plane is physically connectable with the control ports of a VLAN/SAN switch so that it may explain below. Carrying out interconnection of many VLAN switches, and generating complicated multi-yard data networks is known in this field. For example, G. "Designing by Haviland. High-Performance Campus. Refer to available information from Intranets with Multilayer Switching(design of highly efficient yard intranet which has multilayer change)"Cisco Systems, Inc., and Brocade.

[0078]

SAN architecture

SAN is premised on a fiber channel switch, disk apparatus, and comprising fiber channel edge apparatus, such as a SCSI versus Fibre Channel bridge, potentially in explanation. However, SAN may be constituted using the switch which uses other art, such as a Gigabit Ethernet (registered trademark) switch, or other PHYsical layer protocols. The trial in which SAN will be built on an IP network is performed by performing a SCSI protocol on IP especially. The method and the architecture which were mentioned above can be adapted for other SAN constructing methods of these. When building SAN by performing protocols, such as SCSI, on IP in VLAN possible layer 2 environment, a SAN zone is generated by mapping these in different VLAN.

[0079]

The network attaching DOSUTO rage (NAS) which operates on LAN art, such as high-speed Ethernet (registered trademark) or Gigabit Ethernet (registered trademark), may be used. By this choice, in order to strengthen logic partitioning of integrity and a computing grid, different VLAN(s) instead of a SAN zone are used. Such NAS apparatus supports Network File Systems, such as a NSF protocol of Sun, and SMB of Microsoft, and enables it to usually share storage with many same nodes.

[0080]

Operation of control plane

Control plane may be carried out as 1 connected to control and the dataport of SAN and a VLAN switch, or 2 or more processing resources so that it may state here. Various control planes can be carried out and this invention is not restricted to operation of specific control plane. The consideration about the following paragraph 1 control-plane architecture, 2 master-segment manager selection, three controlling functions, four plans, and preservation explains various fields of control plane operation in detail.

[0081]

1. Control plane architecture

According to one example, control plane is carried out as a controlling process hierarchy. The controlling process hierarchy contains 1 or two or more master segment manager mechanisms which a communication interface is carried out to 1 or two or more slave segment manager mechanisms, and generally control these. One or two or more slave segment manager mechanisms control 1 or two or more firm managers. One or two or more firm managers manage 1 or two or more VSF(s). A master and a slave segment manager mechanism may be carried out in hardware circuitry, computer software, or which combination.

[0082]

Drawing 9 is the block diagram 900 showing the logical relation between the control plane 902 and the computing grid 904 by one example. The control plane 902 via the special control port or interface of the networking in the computing grid 904, and a storage element, Computing, the networking, and the storage element which are contained in the computing grid 904 are controlled and managed. The computing grid 904 contains much VSF906 or the logic resource group generated by the example mentioned above.

[0083]

According to one example, the control plane 902 contains the master segment managers 908 and 1, the two or more slave segment managers 910 and 1, or the two or more firm managers 912. The master segment manager 908, the slave segment manager 910, and the firm manager 912 may be stationed at the same position on a specific computing plat form, or may be distributed on many computing plat forms. Although only the single master segment manager 908 is illustrated and explained for convenience, many master segment managers 908 may be used.

[0084]

The communication interface of the master segment manager 908 was carried out to the slave segment manager 910, and he has controlled and managed this. The communication interface of each slave segment manager 910 is carried out to 1 or the two or more firm managers 912, and he manages this. According to one example, each firm manager 912 is stationed at the same position on the computing plat form same as the corresponding slave segment manager 910 by which the communication interface was carried out. The firm manager 912 establishes, constitutes and maintains VSF906 on the computing grid 904. According to one example, single VSF906 which each firm manager 912 manages is assigned, but also for the firm manager 912, much VSF906 are assigned. The firm manager 912 communicates via each slave segment manager 910 rather than is direct respectively. The slave segment manager 910 supervises the assigned condition of the firm manager 912. The slave segment manager 910 makes a stall and the firm manager 912 who did abnormal termination and who was

assigned, respectively resume.

[0085]

The master segment manager 908 supervises loading of VSF906, and determines the quantity of the resource assigned to each VSF906. The master segment manager 908 responds at the time of necessity, passes the firm manager 912, and directs to assign and assign and to cancel the resource of VSF to the slave segment manager 910. According to the necessary condition of specific application, various load-balancing algorithms may be carried out and this invention is not limited to the specific load-balancing method.

[0086]

The master segment manager 908 supervises the loading information on a computing plat form that the slave segment manager 910 and the firm manager 912 are performed, It is judged whether the computing grid 904 is served appropriately. The master segment manager 908 performs the slave segment manager's 910 assignment and quota release, In order to manage the computing grid 904 appropriately if needed, it is directed that the slave segment manager 910 performs the firm manager's 912 assignment and quota release. In order that the master segment manager 908 may also make load balance between the firm manager 912 and the slave segment manager 910 if needed according to one example, Assignment of VSF to the firm manager 912 and the firm manager's 912 assignment to the slave segment manager 910 are managed. According to one example, the slave segment manager 910 communicates as actively as the master segment manager 908, and performs a demand of the change request to the computing grid 904 and another slave segment manager 910, and/or the firm manager 912. When the processing plat form which is performing 1, the two or more slave segment managers 910 and 1, or the two or more firm managers 912 stops functioning, The master segment manager 908 re-assigns other firm managers 912 VSF906 from the firm manager 912 of the stopped computing plat form. In this case, the master segment manager 908 can also direct to start another firm manager 912 to the slave segment manager 910, in order to perform re-assignment of VSF906. Overall power consumption is controllable by managing actively the computing resources of a large number assigned to VSF906, many active firm managers 912, and the slave segment manager 910. For example, shut [the master segment manager 908 / the computing plat form which does not have the active slave segment manager 910 or the firm manager 912] in order to save electric power. Power saving becomes important at the big computing grid 904 and the control plane 902.

[0087]

According to one example, the master segment manager 908 manages the slave segment manager 910 by using registry. Registry includes the information about the present slave segment managers 910, such as the state, the assigned firm manager 912, and VSF906 which were assigned. If the slave segment manager 910 is assigned and quota canceled, registry will

be updated and the slave segment manager's 910 change will be reflected. For example, if the new slave segment manager 910 is illustration-ized by the master segment manager 908 and 1 assigned, or two or more VSF906, Registry is updated and the new slave segment manager 910, and its assigned firm manager 912 and generation of VSF906 are reflected. Next, the master segment manager 908 can judge how it is good to investigate registry periodically and to assign the slave segment manager 910 VSF906.

[0088]

According to one example, registry includes the information about the master segment manager 908 which the master segment manager 910 can access. For example, since registry may contain the data which identifies 1 or the two or more active master segment managers 908, If the new slave segment manager 910 is generated, the new slave segment manager 910 can check registry, and can check about discernment of 1 or the two or more master segment managers 908.

[0089]

Registry may be carried out in various forms and this invention is not limited to a specific practice. For example, registry may be a data file saved in the database 914 in the control plane 902. Registry does not need to be saved out of the control plane 902. For example, registry may be saved at the memory storage of the computing grid 904. In this example, memory storage serves as control plane 902 exclusive use, and is not assigned to VSF906.

[0090]

2. Master segment manager election

Generally, a master segment manager is elected, after the existing master segment manager breaks down when control plane is established or. Although a control plane pair is carried out and a single master segment manager exists, elect two or more master segment managers, and carry out [that it may generally be / the more specific one / advantageous] management by synchronization of the slave segment manager of control plane.

[0091]

According to one example, the slave segment manager in control plane elects the master segment manager of the control plane. There is no master segment manager, and in the simple case where only a single slave segment manager exists, a slave segment manager turns into a master segment manager, and assigns another slave segment manager if needed. When two or more slave segment managers exist, two or more slave processes elect a new master segment manager by the vote of a quorum etc.

[0092]

Since the slave segment manager of control plane is not necessarily permanent, he may choose a specific slave segment manager and may make it participate in a vote. For example, according to one example, the register contains each slave segment manager's time stamp

periodically updated by each slave segment manager. The slave segment manager who was determined in accordance with the specified selection criterion and who has the time stamp updated most these days is considered to still perform, and he is chosen in order to elect a new master segment manager. For example, the newest slave segment manager of the number of specification may be chosen as a vote.

[0093]

According to one example, an election sequence number is assigned to all the active slave segment managers, and a new master segment manager is determined based on an active slave segment manager's election sequence number. For example, the lowest or highest election sequence number may be used, and a specific slave segment manager may be chosen as the next master segment manager (or beginning).

[0094]

When a master segment manager is established, the slave segment manager of the same control plane as a master segment manager, By contacting the present master segment manager (ping), a master segment manager is examined periodically and it is judged whether a master segment manager is still active. When it is judged that the present master segment manager is not active, a new master segment manager is elected.

[0095]

Drawing 10 shows the constitutional diagram 1000 of the master segment manager election by an example. A slave segment manager waits for the end of a ping timer in the state 1002 of being a slave segment manager's main loop. After a ping timer is completed, it will be in the state 1004. In the state 1004, a slave segment manager does the ping of the master segment manager. In the state 1004, a slave segment manager's time stamp (TS) is updated. When a master segment manager answers a ping, the master segment manager is still active and returns to the state 1002. If specific time after does not have a master segment manager to a response, it will be in the state 1006.

[0096]

In the state 1006, the list of active slave segment managers is obtained, and it will be in the state 1008. In the state 1008, it is checked whether other slave segment managers have received the response from a master segment manager. Instead of sending a message to a slave segment manager, in order to perform this check, this information is acquired from a database. . A slave segment manager does not agree for a master segment manager not to be active. That is, when 1 or two or more slave segment managers receive a timely response from a master segment manager, it is presumed that the present master segment manager is still active, and it returns to the state 1002. the "case where a specific number of slave segment managers do not receive a timely response from the present master segment manager -- the present master segment manager -- it is dead" -- that is, it is presumed that it is

not active and it progresses to the state 1010.

[0097]

In the state 1010, the slave segment manager who started the process searches the present election number and the election number of a database to the next from an election table. Next, a slave segment manager updates an election table and writes the entry which specifies the following election number and the most important address in a master election table. Next, a slave segment manager progresses to the state 1012 of reading the lowest sequence number of the present election number. In the state 1014, a specific slave segment manager checks whether it has the lowest sequence number. When not had, it returns to the state 1002. When had, a specific slave segment manager progresses to the state 1016 of becoming a master segment manager. Next, it progresses to the state 1018 and an election number is **********ed.

[0098]

As mentioned above, generally a slave segment manager assigns new VSF according to service of the assigned VSF, and the command from a master segment manager. A slave segment manager also performs a new master segment manager's election with a master segment manager's check again if needed.

[0099]

Drawing 11 is the constitutional diagram 1100 showing various conditions of the slave segment manager by an example. Processing starts in the slave segment manager start state 1102. From the state 1102, it progresses to the state 1104 according to the demand which checks the present master segment manager's condition. In the state 1104, a slave segment manager sends a ping to the present master segment manager, and judges whether the present master segment manager is still active. If there is a timely response from the present master segment manager, it will progress to the state 1106. In the state 1106, the simultaneous transmissive communication of the message is carried out to other slave segment managers, and it tells that the master segment manager answered the ping. It returns from the state 1106 to the start state 1102.

[0100]

In the state 1104, if there is no timely master response, it will progress to the state 1108. In the state 1108, the simultaneous transmissive communication of the message is carried out to other slave segment managers, and it tells that a master segment manager did not answer a ping. Next, it returns to the start state 1102. When a sufficient number of slave segment managers incidentally do not receive a response from the present master segment manager, a new master segment manager is elected as mentioned above.

[0101]

If the demand which resumes VSF from a master segment manager is received from the state

1102, it will progress to the state 1110. In the state 1110, VSF is resumed and it returns to the start state 1102.

[0102]

As mentioned above, VSF of the computing grid by which a master segment manager generally controls a master segment manager is appropriately served by 1 or two or more slave segment managers. For this reason, a master segment manager performs a periodical medical checkup of all the slave segment managers of the same control plane as a master segment manager. According to one example, the master segment manager 908 demands state information periodically from the slave segment manager 910. Information includes which VSF906 is served by the slave segment manager 910, for example. If the specific slave segment manager 910 does not answer within specific time, the master segment manager 908 tries the specific slave segment manager's 910 resumption. When the specific slave segment manager 910 cannot be resumed, the master segment manager 908 re-assigns the firm manager 912 to another slave segment manager 910 from the abnormal slave segment manager 910. Next, the master segment manager 908 can illustration-ize 1 or two or more another slave segment managers 910, and can perform re-balancing of process loading. According to one example, the master segment manager 908 supervises the state of a computing plat form where the slave segment manager 910 is performed. If abnormalities are in a computing plat form, the master segment manager 908 will assign VSF assigned to the firm manager 912 on an abnormal computing plat form to another computing plat form.

[0103]

Drawing 12 is a master segment manager's constitutional diagram 1200. Processing is started in the master segment manager start state 1202. When the master segment manager 908 performs a periodic medical checkup of the slave segment manager 910 of the controlling surface 902 or this is required from the state 1202, it progresses to the state 1204. As all the slave segment managers 910 predicted [state / 1204], when it answers, it returns to the state 1202. This is produced when all the slave segment managers 910 provide the master segment manager 908 with the specific information which shows that all the slave segment managers 910 are operating ordinarily. Or 1 or the two or more slave segment managers 910 did not answer, when the response which shows 1 or the two or more slave segment managers 910 that it was abnormal is carried out, it progresses to the state 1206.

[0104]

In the state 1206, the master segment manager 908 tries the abnormal slave segment manager's 910 resumption. This can be performed by some methods. For example, the master segment manager 908 does not have a response, or can send a resumption message to the abnormal slave segment manager 910. As all the slave segment managers 910 expected, response, i.e., when resumed satisfactorily, it returns from the state 1206 state 1202. For

example, if the abnormal slave segment manager 910 resumes satisfactorily, the slave segment manager 910 will send a resume acknowledge message to the master segment manager 908. When the states 1206-1 or two or more slave segment managers are not able to be resumed, it progresses to the state 1208. This is produced when the master segment manager 908 does not receive a resume acknowledge message from the specific slave segment manager 910.

[0105]

The master segment manager 908 opts for the present loading of the machine which performs the slave segment manager 910 in the state 1208. In order to acquire the slave segment manager's 908 loading information, the master segment manager 908 polls the slave segment manager 910 directly, or acquires loading information from somewhere else, such as the database 914. This invention is not limited to a specific method for the master segment manager 908 to acquire the slave segment manager's 910 loading information.

[0106]

Next, it progresses to the state 1210 and VSF906 assigned to the abnormal slave segment manager 910 is re-assigned to other slave segment managers 910. The slave segment manager 910 to whom VSF906 is assigned informs the master segment manager 908 of when re-assignment was completed. For example, it can tell that the slave segment manager 910 sent the re-quota confirmation message to the master segment manager 908, and re-assignment of VSF906 was completed satisfactorily. It stops at the state 1210 until re-assignment of all the VSF906 relevant to the abnormal slave segment manager 910 is checked. If checked, it will return to the state 1202.

[0107]

Instead of re-assigning other active slave segment managers 910 VSF906 relevant to the abnormal slave segment manager 910, The master segment manager 908 may assign another slave segment manager 910, and may assign the new slave segment manager 910 these VSF906. Selection of whether to re-assign the existing slave segment manager 910 or the new slave segment manager 910 VSF906, It depends on the waiting time relevant to the new slave segment manager's 910 assignment, and the waiting time relevant to re-assignment of VSF906 to the existing slave segment manager 910 selectively at least. Any method can be used according to the necessary condition of specific application, and this invention is limited to neither of the methods.

[0108]

3. Controlling function

According to one example, the communication interface of the control plane 902 is carried out to the global grid manager. The controlling surface 902 provides a global grid manager with fee collection, an obstacle, capacity, loading, and other computing grid information. Drawing 13 is

a block diagram explaining use of the global grid manager by an example.
[0109]
In drawing 13, the partition of the computing grid 1300 is carried out to the logic section called the grid segment 1302. Each grid segment 1302 contains the control plane 902 which controls and manages the data plain 904. In this example, although each data plain 904 is the same as that of the computing grid 904 of drawing 9, It is called a "data plain" in order to explain use of the global grid manager who manages, many the control plane 902 and data plains 904 1302, i.e., grid segments.
[0110]
The communication interface of each grid segment is carried out to the global grid manager 1304. The global grid manager 1304, the control plane 902, and the computing grid 904, Simultaneous arrangement is carried out, or a single computing plat form may be distributed on many computing plat forms, and this invention is not limited to a specific practice.
[0111]
The global grid manager 1304 offers central control of two or more grid segments 1302, and service. The global grid manager 1304 can collect the fee collection from the control plane 902 used by various management tasks, loading, and other information. For example, accounting information is used and the service which the computing grid 904 provides is charged.
[0112]
4. Plan and consideration about preservation
As mentioned above, related VSF in a computing grid and communication must be possible for the slave segment manager in control plane. Similarly, the related slave segment manager and communication must be possible for VSF in a computing grid. VSF in a computing grid does not communicate mutually, in order for a certain VSF to prevent changing the structure of other VSF(s) by a certain method. Various methods of carrying out these plans are explained.
[0113]
Drawing 14 is the block diagram 1400 of the architecture which connects control plane to the computing grid by an example. The control ("CTL") port of the SAN switch (SAN SW1 - SAN SWn) collectively identified with the VLAN switch (VLAN SW1 - VLAN SWn) and the reference number 1404 which are collectively identified with the reference number 1402, It is connected to the Ethernet (registered trademark) subnet 1406. The Ethernet (registered trademark) subnet 1406 is connected to two or more computing elements (CPU1, CPU2 - CPUn) collectively identified with the reference number 1408. Therefore, the communication interface only of the computing element of the control plane 1408 is carried out to the control ports (CTL) of the VLAN switch 1402 and the SAN switch 1404. This structure prevents the computing element in VSF (not shown) changing the membership of VLAN relevant to itself or other VSF(s), and a SAN zone. This method can also be applied when control ports are a

serial or a parallel port. In this case, a port is connected to the computing element of the control plane 1408.

[0114]

Drawing 15 is the block diagram 1500 showing the structure of connecting the control plane computing element (CP CPU1, CP CPU2 - CP CPUn) 1502 by an example to a dataport. In this composition, the control plane computing element 502 sends a packet to the control plane agent 1504 who operates for the control plane computing element 1502 periodically. The control plane agent 1504 polls the computing element 502 periodically for real time data, and sends data to the control plane computing element 1502. The communication interface of each segment manager in the control plane 1502 is carried out to control plane (CP) LAN1506. The communication interface of CP LAN1506 is carried out to the special port V17 of the VLAN switch 504 via the CP firewall 1508. By this structure, an extensible positive means is given to the control plane computing element 1502, and real time information is collected from the computing element 502.

[0115]

Drawing 16 is the block diagram 1600 of the architecture which connects control plane to the computing grid by an example. The control plane 1602 contains control plane computing element CP CPU1, CP CPU2 - CP CPUn. The communication interface of each control plane computing element CP CPU1 in the control plane 1602, CP CPU2 - CP CPUn is carried out to the port S1 of two or more SAN switches which form the SAN mesh 1604 on the whole, S2 - Sn.

[0116]

The SAN mesh 1604 contains the SAN port So and Sp by which a communication interface is carried out to the memory storage 1606 which contains the data which is private life to the control plane 1602. The memory storage 1606 is shown in drawing 16 as a disk for convenience. Memory storage 1606 may be carried out with the storage of which type, and this invention is not limited to the storage of the specific kind of memory storage 1606. The memory storage 1606 is arranged logically in the control plane private memory zone 1608. The control plane private memory zone 1608 maintains the log file, the statistical data, and the present control plane configuration information which carry out control plane 1602. The SAN port So and Sp are the only portions of a control plane private memory zone, and since it is not arranged in other SAN zones, only the computing element in the control plane 1602 can access the memory storage 1606. S1, S2 - Sn, So, and Sp exist in the control plane SAN zone of only a communication interface being carried out to the computing element in the control plane 1602. These ports cannot access the computing element (not shown) in VSF.

[0117]

According to one example, when specific computing element CP CPU1, CP CPU2 - CP CPUn

need to access memory storage or its part, it is a part of specific VSF, and a specific computing element is placed in the SAN zone of specific VSF. For example, computing element CP CPU2 shall access the VSFi disk 1610. In this case, the port s2 relevant to control plane CP CPU2 is arranged in the SAN zone of VSFi containing port Si. Once computing element CP CPU2 accesses to the VSFi disk 1610 of port Si, computing element CP CPU2 will be removed from the SAN zone of VSFi.

[0118]

Similarly, computing element CP CPU1 shall access the VSFj disk 1612. In this case, computing element CP CPU1 is arranged in the SAN zone relevant to VSFj. As a result, the port S1 is arranged in the SAN zone relevant to VSFj which has a zone including the port Sj. Once computing element CP CPU1 accesses to the VSFj disk 1612 connected to the port Sj, computing element CP CPU1 will be removed from the SAN zone relevant to VSFj. The completeness of the control plane computing element by controlling access to a resource by this method correctly using exact SAN zone control and the control plane memory zone 1608 is obtained.

[0119]

As mentioned above, the single control plane computing element can manage two or more VSF(s). Therefore, the single control plane computing element must be able to clarify self in many VSF(s) simultaneously, performing the firewall between VSF(s) in accordance with the plan rule established so much by each control plane. A plan rule may be carried out by preservation or the central segment manager 1302 (drawing 13) in the database 914 (drawing 9) of each control plane.

[0120]

According to one example, since SUPUFU [the VLAN tag based on a port (physical switch)], it combined between VLAN tugging and IP addresses firmly, and has prevented the SUPUFU attack by VSF. The IP packet sent with a certain VLAN interface must have the same VLAN tag and IP address as the logical interface in which a packet arrives. The IP-spoofing attack which the inaccurate server in VSF by this the sauce IP address in another VSF, and changes the logical structure of another VSF potentially, or destroys preservation of a computing grid function is prevented. In the way method of preventing this VLAN tugging, physical access to the computing grid which can be prevented using a Takeshi Takayasu (class A) data center is required.

[0121]

Using various network frame tugging forms, the tag of a data packet may be performed and this invention is not limited to a specific tugging form. According to one example, although other forms are suitable, the VLAN tag of IEE802.1q is used. In this example, VLAN / IP address consistency check is performed with the subsystem of IP stack with which 802.1q tag

information exists, in order to control access. In this example, the computing element comprises a VLAN possible Network Interface Card (NIC) so that the communication interface of the computing element may be simultaneously carried out to many VLAN(s).

[0122]

Drawing 17 is the block diagram 1700 of composition of combining firmly between the VLAN tags and IP addresses by an example. The communication interface of the computing elements 1702 and 1704 is carried out to the ports v1 and v2 of the VLAN switch 1706 via NIC1708 and 1710, respectively. The communication interface also of the VLAN switch 1706 is carried out to the access switches 1712 and 1714. The ports v1 and v2 comprise tag format. According to one example, the VLAN tag information on IEEE802.1q is provided by the VLAN switch 1706.

[0123]

Broader-based computing grid

VSF mentioned above is distributed on WAN by various methods.

[0124]

In one method, broader-based backbone may be based on the Asynchronous Transfer Mode (ATM) change. In this case, each local-area VLAN is ATM. It is extended to a wide area using emu rhe TEDDO LAN (ELAN) which is a part of LAN emulation (LANE) standard. Thus, single VSF spreads to some broader-based whole link, such as ATM/SONET/OC-12 link. ELAN turns into a part of VLAN extended to whole ATM WAN.

[0125]

In other methods, VSF is extended to the whole WAN using a VPN system. In this example, the network fundamental feature becomes unsuitable, carries out interconnection of the two or more VSF(s) over the whole WAN using VPN, and generates the single distribution VSF.

[0126]

In order to carry out the logic copy of the data in the distribution VSF, data mirroring art can be used. Or one of some SAN versus WAN bridging art, such as SAN versus ATM bridging or SAN versus Gigabit Ethernet (registered trademark) bridging, is used, and the bridge of the SAN is carried out on WAN. Since the IP operates satisfactorily on such a network, SAN constituted on the IP network is automatically extended on WAN.

[0127]

Drawing 18 is a block diagram of two or more VSF(s) which are on WAN connection and were extended. The San Jose center, the New York center, and the London center are connected by WAN connection. Each WAN connection comprises ATM, ELAN, or VPN connection, as mentioned above. Each center comprises at least one VSF and at least one idol pool. For example, the San Jose center has VSF1A and the idol pool A. In this composition, the computing resources of each idol pool of a center can be used to the assignment to VSF in

other centers, or specification. If such assignment or specification is performed, VSF will be extended on WAN.

[0128]

The example of use of VSF

The VSF architecture explained in the above-mentioned example may be used by the slag of a web server system. Therefore, the above-mentioned example was explained about the web server, application server, and database server which were constituted from a CPU in specific VSF. However, the VSF architecture may be used in many of other computing situations, and service of other kinds may be provided, and this invention is not limited to a web server system.

[0129]

- Distribution VSF as a part of contents distributed network

In one example, VSF provides a contents distributed network (CDN) using the wide area VSF. CDN is a network of the cash advance server which performs the distributed cash advance of data. The network of a cash advance server can be carried out using the TrafficServer (TS) software currently sold from Inktomi Corporation, San Mateo, and California, for example. TS is a cluster aware system, and if further many CPUs are added to a set of a cash advance traffic server computing element, it will extend a system. Therefore, the addition of CPU is dramatically suitable for the system which is an expansive mechanism.

[0130]

In this composition, since the system can add further many CPUs to the portion of VSF which performs cash advance software, such as TS, dynamically, it can increase cache capacity at the point near the web traffic of the letter of a burst arising. As a result, CDN is constituted so that it may extend dynamically in CPU and I/O bandwidth by a lawful method.

[0131]

- VSF of HOSUTEDDO intranet application

As a host and managed service, the interest to offer of intranet applications, such as a company resource planning (ERP), ORM, and customer-relationship-management software, is increasing. By art, such as Citrix WinFrame and Citrix MetaFrame, a company, Microsoft Windows (registered trademark) application can be provided as service on small lightweight clients, such as Windows (registered trademark) CE apparatus or a web browser. VSF can be acted as a host of such application extensible.

[0132]

For example, the company can make balance load using many applications and data servers with SAP R / 3 ERP software currently sold from SAP Aktiengesellschaft of Germany. In order to extend VSF based on a demand of real time or other factors in VSF, a company adds dynamically further many application servers (for example, SAP dialog server) to VSF.

[0133]
the same -- Citrix Metaframe -- the server farm top which performs HOSUTEDDO Windows (registered trademark) application by adding many Citrix servers -- Windows (registered trademark) -- an application -- a SHON user is extensible. In this case, to VSF, Citrix MetaFrame VSF adds further many Citrix servers dynamically, in order to accommodate the user of the Windows (registered trademark) application in which further many Metaframe(s) act as a host. It becomes clear to act as a host like the example which many of other applications mentioned above.
[0134]
- A customer interaction with VSF
The VSF customer or organization "owns" VSF since VSF responds for asking and is generated can influence each other with a system by various methods, in order to customize VSF. For example, since VSF passes control plane and is generated and changed immediately, privilege access is allowed, and a VSF customer may generate and change the VSF itself. Privilege access is given using a web page and preservation application, token card attestation, the Kerberos exchange, or the pass word authentication given by other suitable preservation elements.
[0135]
The web page of one set is supplied by a computing element or the separate server in one example. The number of computing elements [in / by a web page / in a customer / the number of layers, and a specific layer], The hardware and the software platform which are used to each element, which kind of web server, an application server, or database server software -- custom-made VSF is generable by specifying whether it constitutes from on the computing element of these a priori. Therefore, the customer has the virtual supply console.
[0136]
After a customer or a user inputs such feed information, control plane analyzes and evaluates an order, and in order to perform it, it puts it into queuing. Human being's administrator can re-evaluate an order and a suitable thing can be checked. A credit check of a company is performed and it can check having a suitable credit which makes payment to the demanded service. If a supply order is recognized, control plane will constitute VSF which suits an order and will return to a customer the password which gives the route access to 1 or two or more computing elements in VSF. Next, the customer can upload the master copy of application and can perform by VSF.
[0137]
When the company which adopts a computing grid is a company aimed at obtaining profit, the information about payments, such as a credit card, PO number, an electronic check, or other payment methods, can also be received from a web page.

[0138]
In another example, the customer can choose one of some VSF service plans, such as automatic scaling of VSF between the minimum number of an element, and the maximum number, by a web page based on real-time loading. The customer can have a control value which enables change of parameters, such as a period which must have a minimum number of the computing element in specific layers, such as a web server, or the VSF minimum server capacity. The parameter may be linked to the billing software which adjusts a customer's bill rate automatically and generates a fee collection log file item.
[0139]
The customer can get a report, can supervise the real time information about the hit count or the number of transactions of use, loading, and per second, and can adjust the feature of VSF based on real time information with a privilege access mechanism. The advantage which was superior to the method by the conventional hand control over construction of a server farm according to the above-mentioned special feature is acquired. In the conventional method, the user cannot add a server by various methods and cannot change the characteristic of a server farm automatically, without passing the troublesome manual procedure which constitutes a server farm.
[0140]
- The fee collection model to VSF
Considering the dynamic property of VSF, the company which adopts a computing grid and VSF can use the fee collection model of VSF based on actual use of the computing element of VSF, and a storage element, and can ask for courtesy rates the customer who owns VSF. Since the resource of a certain VSF is not specified statically, the VSF architecture and the method of indicating here make "cash payment" fee collection model possible. Therefore, since the fee relevant to peak server capacity with a specific constant customer whose operating load of the server farm is very changeable is not charged but the fee reflecting the execution average of use, moment use, etc. is charged, a fee can be saved.
[0141]
For example, since it manages using the fee collection model which specifies the time of a company specifying the flat rate to the minimum number of computing elements, such as ten sets of servers etc., and the load of real time needing ten or more elements, For a user, they are charged by the surcharge of an additional server required [how many sets of additional servers] based on the required time. The unit of such fee collection may reflect the resource charged. For example, fee collection may be expressed with units, such as MIPS time, CPU time, and CPU 1000 etc. seconds etc.
[0142]
- Customer visible control plane API

The capacity of VSF is giving a customer the application program interface (API) which specifies the call of the control plane for resource change, and may be controlled by other methods. Therefore, the application program which the customer prepared can emit a call or a demand using API, and also can require many servers and also much storage, still higher throughput, etc. A customer learns this method about computing grid environment, and in order to use the capability which control plane gives, when an application program is needed, it may be used.

[0143]

In neither of the portions, in the above-mentioned architecture, a customer needs to change the application by use with a computing grid. The existing application operates the same with operating in the server farm which carried out manual composition. However, if you understand better the computing resources needed based on the real-time load monitoring function given by control plane, the application can use possible dynamism by a computing grid. If the existing manual method for construction of a server farm is used for API of the above-mentioned character which enables change of the computing capacity of the server farm by an application program, it is not possible.

[0144]

- Automatic updating and versioning

The method and mechanism which are indicated here can be used and the control plane can perform automatic updating and versioning of operating system software which are performed with the computing element of VSF. Therefore, an end user or the customer does not need to be worried about updating an operating system by a new patch, a bug fix, etc. The control plane can maintain the library, if such a software element is received, and it can distribute and install these in the computing element of all the influential VSF(s) automatically.

[0145]

Operation mechanism

A computing element and control plane may be carried out in the form of some, and this invention is not limited to a specific form. In one example, each computing element, It is a general purpose digital computer which has an element shown in drawing 19 except for the nonvolatile storage 1910, and control plane is a general purpose digital computer of the kind shown in drawing 19 which operates under control of the program instruction which carries out the above-mentioned process.

[0146]

Drawing 19 is a block diagram showing the computer system 1900 by which the example of this invention is carried out, and in which it deals. The computer system 1900 contains the processor 1904 connected to the bus 1902, in order to process the bus 1902 or other transmitter styles which transmit information, and information. Since the command which

information and the processor 1904 execute again is saved, the computer system 1900 contains the main memory 1906, such as random access memory (RAM) or other dynamic storage which were connected to the bus 1902. The main memory 1906 can also be used for saving a temporary variable and other intermediate information during execution of the command which the processor 1904 executes. Further, since the command of static information and the processor 1904 is saved, the computer system 1900 contains the read-only memory (ROM) 1908 and other static storages which were connected to the bus 1902. The memory storage 1910, such as a magnetic disk and an optical disc, is formed, and since information and a command are saved, it is connected to the bus 1902.

[0147]

The computer system 1900 may be connected to the display 1912 of a cathode-ray tube (CRT) etc. via the bus 1902, in order to display information on a computer user. The input device 1914 containing an alphanumeric character and other keys is connected to the bus 1902 in order to transmit selection of information and a command to the processor 1904. The user input apparatus of other kinds is the cursor control 1916 of the mouse for transmitting selection of direction information and a command to the processor 1904, and controlling a motion of cursor on the display 1912, a trackball, a cursor arrow key, etc. This input device has two flexibility in two axes which generally enable apparatus to specify the position in a flat surface, i.e., the 1st axis, (for example, x), and the 2nd axis (for example, y).

[0148]

This invention relates to the use of the computer system 1900 for controlling an extensible computing system. According to one example of this invention, control of an extensible computing system is performed by the computer system 1900 according to the processor 1904 which performs 1, 1 of two or more commands, or two or more sequences which are included in the main memory 1906. Such a command is read into the main memory 1906 from the medium which can be read by another computers, such as the memory storage 1910. By performing the sequence of the command included in the main memory 1906, the processor 1904 performs the above-mentioned process process. In multi processing composition, 1 or two or more processors may be used, and the sequence of the command included in the main memory 1906 may be performed. In another example, the circuit by which wiring connection was made may be used combining this instead of a software instruction, and this invention may be carried out. Therefore, the example of this invention is not limited to hardware circuitry and the specific combination of software.

[0149]

The term "medium in which reading in a computer is possible" used here means the medium relevant to giving and executing a command to the processor 1904. Although such a medium contains a nonvolatile medium and volatility medium and a transmission medium, it can take

much form which is not limited to these. A nonvolatile medium contains light or magnetic disks, such as the memory storage 1910. A volatile medium contains dynamic memories, such as the main memory 1906. A transmission medium contains a coaxial cable, copper wire, and optical fiber including the wiring which constitutes the bus 1902. A transmission medium can also take the form of a sound wave which is generated between radio and infrared ray data communication, or a light wave.

[0150]

The general form of the medium which can be read by computer, For example, a floppy (registered trademark) disk, a flexible disk which are explained below, A hard disk, magnetic tape, other magnetic media, CD-ROM, other optical media, Other media which a punch card, a paper streamer, other physical media that have a pattern of a hole, RAM, PROM, EPROM, FLASH-EPROM, other memory chips or a cartridge, a subcarrier, or a computer can read are included.

[0151]

Various forms of the medium which a computer can read may relate to making the processor 1904 send and perform 1, 1 of two or more commands, or two or more sequences. For example, a command is first sent to the magnetic disk of a remote computer. A remote computer loads a command to the dynamic memory, and sends a command on a telephone line using a modem. The modem which is in remoteness to the computer system 1900 can receive the data on a telephone line, and can change data into an infrared signal using an infrared transmitter. The infrared detector connected to the bus 1902 receives the data carried with an infrared signal, and takes out data to the bus 1902. The bus 1902 sends data to the main memory 1906, and the processor 1904 performs search and execution of a command from here. The command which the main memory 1906 received can be saved [in front of execution of the processor 1904, or in the back] optionally at the memory storage 1910.

[0152]

The computer system 1900 also includes the communication interface 1918 connected to the bus 1902. The communication interface 1918 performs bidirectional data communication linked to the network link 1920 connected to the local network 1922. For example, the communication interface 1918 may be the digital synthesis service network (ISDN) card or modem for performing data communication connection to the corresponding telephone line of a kind. In other examples, the communication interface 1918 may be a Local Area Network (LAN) for performing data communication connection to compatible LAN. A radio link can also be carried out. In such operation, the communication interface 1918 transmits and receives the electrical and electric equipment, the electromagnetism, or the lightwave signal which tells the digital data stream showing various kinds of information.

[0153]

Generally the network link 1920 performs the data communications to other data facilities via 1 or two or more networks. For example, the network link 1920 provides connection with the host computer 1924 or data facility managed by Internet Service Provider (ISP) 1926 via the local network 1922. ISP1926 provides data transmission services via the worldwide scale packet data communication network 1928 generally called the "Internet" and now. Both the local network 1922 and the Internet 1928 use the electrical and electric equipment, the electromagnetism, or the lightwave signal which tells a digital data stream. The signal which sends and receives digital data to the computer system 1900 via various networks, the signal on the network link 1920, and the communication interface 1918 is a typical form of the subcarrier which carries information.

[0154]

The computer system 1900 can receive the data which transmits a message and contains a program code via a network, the network link 1920, and the communication interface 1918. In the example of the Internet, the server 1930 transmits the request code of an application program via the Internet 1928, ISP1926, the local network 1922, and the communication interface 1918. According to this invention, such downloaded application specifies control of the extensible computing system explained here.

[0155]

A receiving cord may be saved at the memory storage 1910 or other nonvolatile storage, in order to perform by the processor 1904 in execution and/or the back, if received. Thus, the computer system 1900 can obtain application codes in the form of a subcarrier.

[0156]

The computing grid indicated here is notionally compared with the public power network sometimes called a power grid. A power grid provides much authorized personnel with an extensible means, in order to obtain electric power service via a single large-scale electric power infrastructure. Similarly, the computing grid indicated here provides computing service for many organizations by using a single large-scale computing infrastructure. Since a power grid is used, a power consumption person does not manage the individual power equipment independently. For example, a utility consumption person makes a personal dynamo operate in the equipment or share equipment, and there is no reason for managing the capacity and increase individually. Instead, since the power grid can supply electric power broadly to the great portion of population, big economy of scale is obtained. Similarly, a large-scale single computing infrastructure can be used for the computing grid indicated here, and it can provide computing service for the great portion of population.

[0157]

In the above-mentioned detailed description, this invention was explained in relation to the concrete example. However, it will become clear that it is possible to add various improvement

and change to this invention, without deviating from the vast pneuma and range of this invention. Therefore, explanation and a drawing are taken into consideration not in a restrictive meaning but in illustration.

[Brief Description of the Drawings]

[Drawing 1 A]

Drawing 1 A is a block diagram of the simple website which uses single computing element topology.

[Drawing 1 B]

Drawing 1 B is a block diagram of an one-layer web server firm.

[Drawing 1 C]

Drawing 1 C is a block diagram of a three-layer Web server firm.

[Drawing 2]

Drawing 2 is a block diagram showing one composition of the extensible computing system 200 containing a local computing grid.

[Drawing 3]

Drawing 3 is a block diagram of the typical virtual server farm by which a SAN zone is characterized.

[Drawing 4 A]

Drawing 4 A is a block diagram showing the continuous process relevant to the addition of a computing element, and removal of the element from a virtual server farm.

[Drawing 4 B]

Drawing 4 B is a block diagram showing the continuous process relevant to the addition of a computing element, and removal of the element from a virtual server farm.

[Drawing 4 C]

Drawing 4 C is a block diagram showing the continuous process relevant to the addition of a computing element, and removal of the element from a virtual server farm.

[Drawing 4 D]

Drawing 4 D is a block diagram showing the continuous process relevant to the addition of a computing element, and removal of the element from a virtual server farm.

[Drawing 5]

Drawing 5 is a block diagram of the example of a virtual server farm system, a computing grid, and surveillance.

[Drawing 6]

Drawing 6 is a block diagram of the logical connection of a virtual server farm.

[Drawing 7]

Drawing 7 is a block diagram of the logical connection of a virtual server farm.

[Drawing 8]

Drawing 8 is a block diagram of the logical connection of a virtual server farm.

[Drawing 9]

Drawing 9 is a block diagram of the logical relation of control plane and a data plain.

[Drawing 10]

Drawing 10 is a constitutional diagram of a master control selection process.

[Drawing 11]

Drawing 11 is a constitutional diagram of a slave controlling process.

[Drawing 12]

Drawing 12 is a constitutional diagram of a master controlling process.

[Drawing 13]

Drawing 13 is the control plane of a CC processor and a large number, and a block diagram of a computing grid.

[Drawing 14]

Drawing 14 is a block diagram of the architecture which carries out the portions of control plane and a computing grid.

[Drawing 15]

Drawing 15 is a block diagram of the system which has a computing grid protected by the firewall.

[Drawing 16]

Drawing 16 is a block diagram of the architecture which connects control plane to a computing grid.

[Drawing 17]

Drawing 17 is a block diagram of the arrangement which combines a VLAN tag and an IP address densely.

[Drawing 18]

Drawing 18 is a block diagram of two or more VSF(s) which are on WAN connection and were extended.

[Drawing 19]

Drawing 19 is a block diagram of a computer system in which an example is carried out.

[Translation done.]

CLAIMS

[Claim(s)]
[Claim 1]
A master control mechanism,
A communication interface is carried out to a master control mechanism, and it responds to 1 or two or more commands from a master control mechanism,
The 1st subset of processing resources is chosen from a set of processing resources,
choosing the 1st subset of a memory resource from a set of a memory resource -- and
The communication interface of the 1st subset of processing resources is carried out to the 1st subset of a memory resource,
A control device possessing 1 or two or more slave control mechanisms which were constituted so that the 1st logic resource group containing the 1st subset of processing resources and the 1st subset of a memory resource might be established.
[Claim 2]
The control device according to claim 1 which a master control mechanism is a master controlling process performed on 1 or two or more processors, and is 1 or two or more slave processes that 1 or two or more slave control mechanisms are performed on 1 or two or more processors.
[Claim 3]
The control device according to claim 1 whose master control mechanism is 1 or two or more master processors and 1 or whose two or more slave control mechanisms are 1 or two or more slave processors.
[Claim 4]
A master control mechanism based on loading of a slave controlling process mechanism between 1 or two or more slave control mechanisms, The control device according to claim 1 constituted so that control of 1 or two or more memory resources from a subset of 1 or two or

more processing resources from a subset of processing resources, and a memory resource may be re-assigned dynamically.

[Claim 5]

A master control mechanism based on loading of a slave controlling process mechanism, Assign dynamically 1 or a slave control mechanism of two or more additions, and and control of 1 or two or more memory resources from a subset of 1 or two or more processing resources from a subset of processing resources, and a memory resource, The control device according to claim 1 constituted so that it may assign 1 or two or more slave control mechanisms which were added.

[Claim 6]

A master control mechanism based on loading of a slave controlling process mechanism, . Are already assigned to 1 or two or more specific slave control mechanisms from 1 or two or more slave control mechanisms. re-assigning control of 1 or two or more specific memory resources from a subset of 1 or two or more specific processing resources from a subset of processing resources, and a memory resource to 1, 1 from two or more slave control mechanisms, or other two or more slave control mechanisms -- and

The control device according to claim 1 constituted so that quota release of 1 or two or more specific slave control mechanisms may be performed dynamically.

[Claim 7]

A master control mechanism,

A state of 1 or two or more slave control mechanisms is determined,

trying a reboot of 1 or two or more specific slave control mechanisms, when 1, 1 from two or more slave control mechanisms, or two or more specific slave control mechanisms do not answer correctly or it is not functioning -- and

When 1 or two or more specific slave control mechanisms cannot reboot,

starting 1 or two or more new slave control mechanisms -- and

The control device according to claim 1 which comprises 1 or two or more specific slave control mechanisms so that control of processing resources and a memory resource may be re-assigned to 1 or two or more new slave control mechanisms.

[Claim 8]

One or two or more slave control mechanisms,

determining a state of a master control mechanism -- and

When a master control mechanism carries out abnormal termination or it is not functioning appropriately any longer,

The control device according to claim 1 constituted so that a new master control mechanism may be chosen from 1 or two or more slave control mechanisms.

[Claim 9]

The control device according to claim 1 with which 1 or two or more commands from a master control mechanism are generated based on processing and a memory necessary condition that the 1st logic resource group was expected.

[Claim 10]

One or two or more slave control mechanisms respond to 1 or two or more commands from a master control mechanism further,

Dynamic change of the number of processing resources of the 1st subset of processing resources,

Dynamic change of the number of memory resources of the 1st subset of a memory resource,

In order to make change of the number of processing resources of the 1st subset of processing resources, and the number of memory resources of the 1st subset of a memory resource reflect, The control device according to claim 1 constituted so that a dynamic change of a communication interface between the 1st subset of processing resources and the 1st subset of a memory resource may be made.

[Claim 11]

Change of the number of processing resources of the 1st subset of processing resources and the number of memory resources of the 1st subset of a memory resource, The control device according to claim 10 directed by a master control mechanism based on actual loading of the 1st subset of processing resources, and the 1st subset of a memory resource.

[Claim 12]

One or two or more slave control mechanisms are constituted so that the 2nd logic resource group containing the 2nd subset of processing resources and the 2nd subset of a memory resource may be further established according to 1 or two or more commands from a master control mechanism, and he is the 2nd logic resource group,

The 2nd subset of processing resources is chosen from a set of processing resources,

choosing the 2nd subset of a memory resource from a set of processing resources -- and

The control device according to claim 1 in which communication separation is done by the 1st logic resource group by carrying out the communication interface of the 2nd subset of processing resources to the 2nd subset of a memory resource.

[Claim 13]

The communication interface of the 1st subset of processing resources is carried out to the 1st subset of a memory resource by using 1 or two or more storage area network (SAN) switches,

The communication interface of the 2nd subset of processing resources is carried out to the 2nd subset of a memory resource by using 1 or two or more SAN switches,

The control device according to claim 12 in which communication separation is done by the 1st logic resource group when the 2nd logic resource group uses tugging and SAN zoning.

[Claim 14]

The control device according to claim 13 performed when SAN zoning uses port level SAN zoning or LUN level SAN zoning.

[Claim 15]

The communication interface of the master control mechanism is carried out to a CC mechanism,

a master control mechanism is constituted so that loading information to the 1st logic resource group may be given to a CC mechanism -- and

The control device according to claim 1 constituted so that a master control mechanism may generate 1 or two or more commands to 1 or two or more slave control mechanisms based on 1 or two or more CC commands which were received from a CC mechanism.

[Claim 16]

A process of starting a master control mechanism,

A communication interface is carried out to a master control mechanism, and it responds to 1 or two or more commands from a master control mechanism,

The 1st subset of processing resources is chosen from a set of processing resources,

choosing the 1st subset of a memory resource from a set of a memory resource -- and

The communication interface of the 1st subset of processing resources is carried out to the 1st subset of a memory resource,

How to manage processing resources possessing a process of starting 1 or two or more slave control mechanisms which were constituted so that the 1st logic resource group containing the 1st subset of processing resources and the 1st subset of a memory resource might be established.

[Claim 17]

a process of starting a master control mechanism includes a process of starting a master controlling process performed on 1 or two or more processors -- and

A way according to claim 16 a process of starting 1 or two or more slave control mechanisms includes a process of starting 1 or two or more slave processes which are performed on 1 or two or more processors.

[Claim 18]

a process of starting a master control mechanism includes a process of starting 1 or two or more master control processors -- and

A way according to claim 16 a process of starting 1 or two or more slave control mechanisms includes a process of starting 1 or two or more slave processors.

[Claim 19]

Based on loading of a slave controlling process mechanism, control of 1 or two or more memory resources from a subset of 1 or two or more processing resources from a subset of processing resources, and a memory resource, A method according to claim 16 of including

further a master control mechanism re-assigned dynamically between 1 or two or more slave control mechanisms.

[Claim 20]

Based on loading of a slave controlling process mechanism, 1 or a slave control mechanism of two or more additions is assigned dynamically, And a method according to claim 16 of including further a master control mechanism which assigns control of 1 or two or more memory resources from a subset of 1 or two or more processing resources from a subset of processing resources, and a memory resource to 1 or two or more slave control mechanisms which were added.

[Claim 21]

. Based on loading of a slave controlling process mechanism, are already assigned to 1, 1 from two or more slave control mechanisms, or two or more specific slave control mechanisms. Control of 1 or two or more specific memory resources from a subset of 1 or two or more specific processing resources from a subset of processing resources, and a memory resource, A method according to claim 16 of including a re-assigned master control mechanism in 1, 1 from two or more slave control mechanisms, or other two or more slave control mechanisms further.

[Claim 22]

A state of 1 or two or more slave control mechanisms is determined,

trying so that 1 or two or more specific slave control mechanisms may be rebooted when not functioning appropriately, or 1, 1 from two or more slave control mechanisms, or two or more specific slave control mechanisms do not answer -- and

When 1 or two or more specific slave control mechanisms cannot reboot, 1 or two or more new control mechanisms are started, And a method according to claim 16 of including further a master control mechanism which re-assigns control of processing resources and a memory resource in 1 or two or more new slave control mechanisms from 1 or two or more specific slave control mechanisms.

[Claim 23]

determining a state of a master control mechanism -- and

A method according to claim 16 of including further 1 or two or more slave control mechanisms which choose a new master control mechanism from 1 or two or more slave control mechanisms when a master control mechanism carries out abnormal termination or it is not functioning appropriately any longer.

[Claim 24]

A method according to claim 16 by which 1 or two or more commands from a master control mechanism are generated based on processing and a memory necessary condition that the 1st logic resource group was predicted.

[Claim 25]
It responds to 1 or two or more commands from a master control mechanism,
Dynamic change of the number of processing resources in the 1st subset of processing resources,
Dynamic change of the number of memory resources in the 1st subset of a memory resource,
In order to make change of the number of processing resources in the 1st subset of processing resources, and the number of memory resources in the 1st subset of a memory resource reflect, A method according to claim 16 of including further 1 which makes a dynamic change of a communication interface between the 1st subset of processing resources, and the 1st subset of a memory resource, or two or more slave control mechanisms.
[Claim 26]
Change of the number of processing resources in the 1st subset of processing resources and the number of memory resources in the 1st subset of a memory resource, A method according to claim 25 directed by a master control mechanism based on actual loading of the 1st subset of processing resources, and the 1st subset of a memory resource.
[Claim 27]
The 2nd logic resource group including further 1 which establishes the 2nd logic resource group containing the 2nd subset of processing resources, and the 2nd subset of a memory resource according to 1 or two or more commands from a master control mechanism, or two or more slave control mechanisms,
Selection of the 2nd subset of processing resources from a set of processing resources,
selection of the 2nd subset of a memory resource from a set of processing resources -- and
A method according to claim 16 in which signal transduction separation is carried out from the 1st logic resource group by communication interface of the 2nd subset of processing resources to the 2nd subset of a memory resource.
[Claim 28]
The communication interface of the 1st subset of processing resources is carried out to the 1st subset of a memory resource by using 1 or two or more storage area network (SAN) switches,
the communication interface of the 2nd subset of processing resources is carried out to the 2nd subset of a memory resource by using 1 or two or more SAN switches -- and
A method according to claim 27 by which communication separation of the 2nd logic resource group is done by the 1st logic resource group tugging and by carrying out SAN zoning.
[Claim 29]
A method according to claim 28 performed when SAN zoning uses port level SAN zoning or LUN level SAN zoning.
[Claim 30]
The communication interface of the master control mechanism is carried out to a CC

mechanism,
A master control mechanism is constituted so that the 1st logic resource group's loading information may be given to a CC mechanism,
A method according to claim 16 constituted so that a master control mechanism may generate 1 or two or more commands of 1 or two or more slave control mechanisms further based on 1 or two or more CC commands which were received from a CC mechanism.
[Claim 31]
If it is a medium which can be read by computer which tells 1 for managing processing resources, 1 of two or more commands, or two or more sequences and 1, 1 of two or more commands, or two or more sequences are performed by 1 or two or more processors, 1 or 2 or more processors,
A process which makes a master control mechanism start,
A communication interface is carried out to a master control mechanism, and according to 1 or two or more commands from a master control mechanism, A process which makes 1 or two or more slave control mechanisms which are constituted so that the 1st logic resource group containing the 1st subset of processing resources and the 1st subset of a memory resource may be established start,
The 1st subset of processing resources is chosen from a set of processing resources,
choosing the 1st subset of a memory resource from a set of a memory resource -- and
A medium which is performed by carrying out the communication interface of the 1st subset of processing resources to the 1st subset of a memory resource and in which reading in a computer is possible.

[Translation done.]